Statistics Canada International Symposium Series - Proceedings

Symposium 2003: Challenges in Survey Taking for the Next Decade

2003
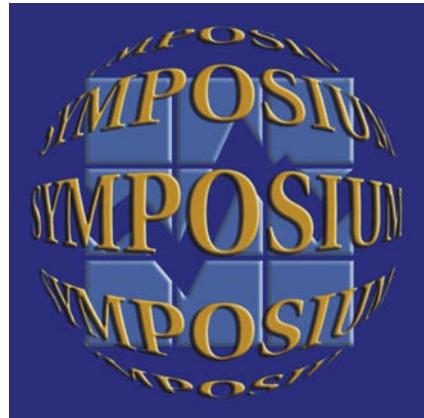
Canada

Proceedings of Statistics Canada Symposium 2003
Challenges in Survey Taking for the Next Decade

MAKING DATA QUALITY VISIBLE IN PRACTICE

Frank Yu[1]

ABSTRACT

To help users of statistical data understand data quality in all its aspects, and access and utilise information about quality, the Australian Bureau of Statistics (ABS) has embarked on a "Making Data Quality Visible" program. In this paper, we will report our experience with increasing the information available internally and externally about the quality of ABS data, and educating our users and ourselves on how knowledge of data quality of particular datasets could help inform the uses of statistics. We will describe how the broad strategy is implemented in practice, challenges and constraints encountered, and solutions to the problems that emerged.

To ensure high quality statistics are produced, monitoring of quality measures should be included as an integral part of the production process as well as the quality assurance process for signing off publications. We have commenced building a "Quality Information Infrastructure" for capturing, storing and making available quality measures/indicators to support both within cycle and end of cycle use. The infrastructure should link seamlessly to the production systems and other corporate facilities for managing of metadata.

We aspire to enhancing our paper and electronic publications by including more information about data quality. As well, users of statistical data need to understand, access and use information about quality to improve the use of particular datasets. To this end, we are developing a suite of courses to help users make appropriate use of the data. In this paper, we will discuss some practical issues relating to making quality more visible in the user community.

KEYWORDS: Education Strategy; Quality Assurance; Quality Management; Statistical Infrastructure.

1. INTRODUCTION

In recent years, making quality visible has become a major focus for National Statistical Offices as well as international agencies. Quality frameworks and guidelines have been developed for the International Monetary Fund (Carson, 2000), Eurostat (2001) and the Organisation for Economic Cooperation and Development (2003). This is by no means coincidental, as leaders of statistical agencies understand that confidence in the quality of the information they produce is a survival issue for statistical agencies (Brackstone, 1999). Fellegi (1996) provides a strong argument that the intrinsic value and useability of information depends directly on the credibility of the statistical system. While few users can validate directly the data released by statistical agencies, and will often rely on the reputation of the provider of the information to judge their quality, the continuation of their trust will depend on the statistical agencies being able to demonstrate the quality and objectivity of their products. Indeed, users often distinguish National Statistical Offices with other providers of data, by their ability to produce data that fit their purpose.

The quality of statistical data is enhanced if users can access, understand and use information about the quality of data. Users need information on quality of statistical data to help assess fitness for use and manage the risks from using the data (Allen, 2001). It is not sufficient just to present information on the quality of the data, or even to explain how to interpret a quality declaration. That information must also be accessible, useable, and inform the users for the purpose of decision-making or research (Lee and Allen, 2001)

To pursue this objective, the Australian Bureau of Statistics (ABS) launched a program aimed at helping users of statistical data to understand the concepts and framework of data quality, access the information which allows them

---

[1]Frank Yu, Australian Bureau of Statistics, P.O. Box 10, Belconnen, ACT 2616, Australia, frank.yu@abs.gov.au

to assess the quality of particular datasets, and use that information about quality effectively to manage the risks arising from data limitations. The program, entitled "Making Quality Visible" (MQV), is comprised of a range of activities across the ABS, many of which are linked to other corporate objectives. It is not a formally resourced project, more a theme woven throughout many strands of the ABS work. The program has the following key objectives:

- Publish and promote guidelines about frameworks about quality;
- Educate our users about quality, and how to use quality information to inform the uses of statistics;
- Improve the useability of technological tools and processes;
- Use information about data quality to manage and improve our statistical processes; and
- Increase the information available externally about the quality of ABS data

The major drivers of this program are commonly experienced in many National Statistical Offices. The ABS is committed to expanding and improving the dissemination of data, with a particular focus on promoting electronic dissemination and a self help approach to transaction. While electronic delivery of information creates opportunities for new and improved access to data, users are also risking the loss of some aspects of metadata such as warning about data limitations. As well, electronic dissemination provides us with the opportunity to present far more detailed information than we have ever done in the past. The National Statistical Office has the obligation to ensure that clients do not misuse the data through over-perception of their accuracy.

Not only has the quantum of statistical data available to users expanded, the nature of data disseminated is also changing. The ABS, like other statistical agencies, is seeking to expand the national statistical service by making better use of non-ABS data, both by integrating them into ABS products, and by branding those official statistics that are of high quality and importance to society. In extending the national statistical service, there is a need to inform users about the quality attributes of the extensions. Appropriate standards for quality and approach to quality assessment of non-ABS data need to be developed to support the initiative. Another category of statistical output different from the traditional products is analytical products, which are typically obtained from a modelling process and involve elaborate transformation of data. Users will need to grasp the quality characteristics of these products in order to understand and use them appropriately.

Users expectations on quality are changing - they are much higher than a decade ago (Trewin, 2002). National Statistical Offices has to accept the rising demands and do their best to improve quality to the expected level. That may not be always possible and the National Statistical Offices must manage users' expectations. Making quality of the information visible and providing good explanations of the strengths and weaknesses of particular datasets is an important element of this process. As well, as National Statistical Offices seek to improve the efficiency of statistical processes through introducing new ways of processing statistics, new organisational arrangements, improved methodology and revamped technology, there is a need to ensure that quality gains are realised by taking advantage of the innovations. As well, users want to be reassured that the fitness of data for their purpose is not being compromised as a result of the changes.

In this paper, we discuss the strategies adopted for the Making Quality Visible program. The program is still work in progress; this paper reports our initial experience. Our focus is on the practical aspects of the strategies - what we aim to develop, how we are doing it, what has worked and what did not. Section 2 describes the development of quality guidelines for different types of statistical collections/products and our exhortation strategy. Section 3 highlights changes aimed at enhancing the use of quality information internally. Section 4 describes efforts to expand the amount of information published. We discuss our experience and further challenges ahead in section 5.

2. PROMOTING GUIDELINES AND FRAMEWORK ABOUT QUALITY

Information about quality is important for consumers of statistical data, who require the information to assess the risks from quality issues and to make informed decision based on the assessment. It is not adequate for the National Statistical Office just to present information on the quality of the data, or even to explain how to interpret the quality measures. The National Statistical Office must provide leadership and assistance to help users with using the information constructively for their purpose (Allen and Lee, 2001). It is important that common terminology and language is used by the National Statistical Office to describe quality. It is also important for the National Statistical Office to develop an education and exhortation strategy to help users with understanding all aspects of data quality. In this section, we report the ABS' experience with publishing quality framework and guidelines, and exhorting users to adopt them in their work.

Although a range of definitions of quality have been used by different organisations, they tend to be broadly similar to each other. There are some differences, typically reflecting the respective context of application in each organisation, but basically they are providing the same message - there is much more to quality than accuracy (Brackstone, 1999). The ABS has adopted the framework used by Statistics Canada for defining quality. The definition is based on six dimensions of quality - relevance, accuracy, timeliness, coherence, accessibility and interpretability.

More recently, the ABS is considering to add a seventh dimension to defining quality - trust in the integrity of statistical source. Users often based their trust of particular datasets on the objectivity, credibility and statistical capability of the source of the data, and depending on the level of their trust they may decide on how fit the data are for their purpose. The issue is particularly relevant for the ABS as it seeks to extend the use of non-ABS data - it is important to distinguish data providers who commit to following sound principles of statistical collection and data management, with those who either do not wish to commit to these principles or have the capability of adopting them.

The ABS is developing, incrementally, a quality manual which will define the quality framework and present all guidelines relating to the collection and use of information on quality. Issues of relevance, coherence, accessibility, interpretability and timeliness generally span all sources of data, but especially for accuracy the approach to quality assessment and description differ depending on whether the datasets are from a traditional statistical survey/census, an administrative data source, a system of accounts, a price index or a complex analytical products:

- For the traditional sample survey/census category there is a reasonably good understanding of the detail, based on working through the standard collection cycle and nominating the various stages and elements at which indicators of accuracy might impact on output quality.

- The ABS have commenced developing a framework to administrative data. We promoted a National Statistical Service Best Practice Guidelines (ABS, 2001) to espouse many of the sound principles of quality management. There is strong link between output quality and the generating processes and enablers. The quality framework is a useful structure for documenting our experience in managing non-ABS data sources. We have organised a series of internal workshops with a view to reviewing this experience and developing the quality manual for such datasets.

- For systems of economic accounts, the ABS has developed, through a project undertaken in 2000 (Zarb, 2001) a detailed and comprehensive elaboration of the framework for measuring, analysing and dissecting the factors that affect the quality of such type of estimates. The framework takes account of the nature of these estimates, which typically comprise of multiple time series which are seasonally adjusted or trended, and analytically related in special ways. Feedback from the users and analysts community suggest that the approach is too elaborate and populating the fully detailed framework regularly would be overkill. The framework is, however, useful for comprehensively assessing and describing, on an as needed basis, the quality of particular segments of accounting systems; for example, it may be applied to the industry productivity measures or satellite account estimates.

- We have no plans at present to develop a quality framework tailored to price indexes although this may be useful to pursue in future.

- For analytical outputs, some guidance on the quality assurance of elaborately transformed data has been in use. The guidelines are based on the six quality dimensions and emphasise making the quality of such products visible by ensuring the analytical processes and assumptions are transparent to users, and providing guidance about valid and invalid use of the estimates.

Allen (2001) describes the education strategy that is being implemented to help users understand the quality framework and apply information concerning quality to assess how well particular datasets will meet their decision making or research purposes. The most tangible element of the strategy is a suite of courses which the ABS is implementing by stages. In sequence, we target senior ABS managers, then a broader range of ABS staff, and finally the staff of other agencies, including those who are producers of non-ABS data and the consumers of data. The education program would be most effective if it is implemented in alignment with ABS' promotion of the national statistical service, and exhortation of other agencies to adopt sound quality assurance and data management practices.

To date, we have developed and run a course, titled 'Quality Informed Decisions', for some 300 ABS staff. The course seeks to answer the often asked question "so what should I (the user) do with the information about data quality?" The contents of this course is described in Allen (2001). The course is accepted as meeting a basic training requirement and have been embedded into other on-going training program for all graduate recruits. Some ABS subject areas have implemented the contents of the course to help assess the requirements of statistical information for policy making. Within the ABS, there has been a strong buy in of the quality framework and no doubt the education strategy has played its part.

We are adapting the course to tailor it to the requirements of other Government agencies, to ensure that attendees will be able to apply what they learn from the course to their work context. Marketing of the course will be aligned with the process of assessing the suitability of non-ABS data for inclusion as main indicators of official statistics.

Further courses to be developed include one on defining data needs and developing data strategies, a course on quality assurance of data from business surveys; and a course of quality assurance of data from administrative source.

3. MAKING QUALITY VISIBLE IN STATISTICAL PROCESSES

Like many other statistical organisations, the ABS is implementing changes to the way it collects information to improve the efficiency and effectiveness of its statistical programs. There are a number of advantages in focussing on using quality information to improve the efficiency of statistical processes and/or the quality of statistical outputs. First, the survey managers are in control of the change process, being able to identify the trade off between costs and different dimensions of quality and inform users of the implications accordingly. Second, the sharpened focus on quality information will enhance the capability of the survey areas for describing the quality of statistical outputs to users and demonstrate the strengths and weaknesses of the particular datasets. Likewise, users who are familiar with the quality framework could provide informed criticism which could lead to constructive partnership in improving quality. Third, the national statistical office is able to demonstrate the usefulness of information about quality and provide leadership to customers of data on how they may use the quality framework to assess the fitness of the statistical information for their purpose.

The ABS has some well established approach to managing the quality of collections. While the approach adopted by collections may vary according to the nature of data and frequency of publication, there are typically four elements in the quality assurance process:

- a process for clearance of the survey output for publication;
- a mechanism for responding to quality issues that emerge prior to publication;

- continuous improvement of the survey process through improvements of methodology, business process and infrastructure;
- an audit of the quality risks of collections through such process as methodology audit and risk assessments.

To facilitate these processes, we aim to expand the routine production of quantifiable indicators of quality, especially those relating to non-sampling errors, throughout the statistical processing cycle. This in turn feeds into development of guidelines, infrastructure and education strategy to support the use of quality information for the quality assurance process.

Our vision is a future where quality measures should be easily available to various workgroups in real time, to enable them to monitor outcomes and continuously improve the survey process, both for the purposes of quality assurance during the cycle as well as for quality review and improvement post the processing cycle.

The ABS has an initial set of quality measures which has been defined for both household and business surveys and are supported by existing processes and infrastructure. These measures are being expanded to cover the whole statistical cycle. Examples include: quality measures for sampling frames, data capture, non-response (and where possible, their impact on estimates) and the impact of non-response and the quality of the time series analysis for producing seasonally adjusted and trend estimates. Standard measures, once defined, are being built into corporate infrastructure for production. The next step is for the quality indicators to be consistently loaded to corporate data warehouse for each cycle of each collection. This would make them corporately accessible, improving the transparency of survey quality, and positioning the ABS for their eventual publication.

Many survey areas in business statistics have established clearance process to assure the statistical output prior to publication but there is wide variation in the coverage and presentation of quality information presented to support the process. To improve the clearance process, the survey methodologists are working together with collection areas to develop best practice guidelines and introduce an education strategy aimed at enhancing the effectiveness of the process. Our experience is that collection areas either produce too much information with inadequate analysis and scrutiny of quality issues, or information skewed to a couple of standard parameters which do not inform on the particular quality problems relating to the survey. Similarly, for household surveys there is a need for strengthening the output validation process to ensure a timely focus on the quality of outputs.

Assembling information on quality is a prerequisite for making such information more accessible to ABS management as well as customers of data. We do not want this to be a heavy impost on collection areas. To this end, we would design the technological solution to support processing and presentation of quality information such that it should be a useful by-product of the data compilation process. This is the driver for the launch of the 'Quality Infrastructure Project'.

The Quality Infrastructure Project is aimed at building a standard repository for capturing, storing and disseminating both within cycle and end of cycle quality measures/indicators. It should be built in a way which enable it to link seamlessly with corporate facilities for statistical processes, and data warehouse for dissemination. The infrastructure will be applicable for a wide range of collections, including business surveys, household surveys and population census. Quality measures produced as by-product of many ABS statistical processes will be able to be stored on the Quality Infrastructure repository. It will enable the quality information to be viewed from corporate wide desktops, and disseminated internally as well as externally, ABS staff will be able to analyse, report and graph quality measures in the repository using on line analytical processing tools.

Figure 1 illustrates the broad design of the Quality Infrastructure. The extraction and archival of quality measures will be defined and supported by the infrastructure through services which can interact with a wide range of platforms with metadata driven functionalities. Quality indicators are corporately defined and managed within the corporate metadata repository - metadata are sourced from existing data warehouse as far as possible, for example, details of collection, data items, input datasets etc. are sourced from relevant parts of the corporate metadata repository. Indeed, the quality infrastructure component is itself part of the metadata repository. The quality measures repository itself is an Oracle database designed using the dimensional data model. A central fact table holds all the quality measures which are described by the dimensions around it. Descriptions of quality measures will be centrally managed by the Methodology Division to ensure standardisation and coherence.

Fig. 1: Broad design of the Quality Infrastructure

(components for the quality infrastructure are shaded, with stage 1 components being in light colour and stage 2 components in heavier shade.)

Quality Infrastructure
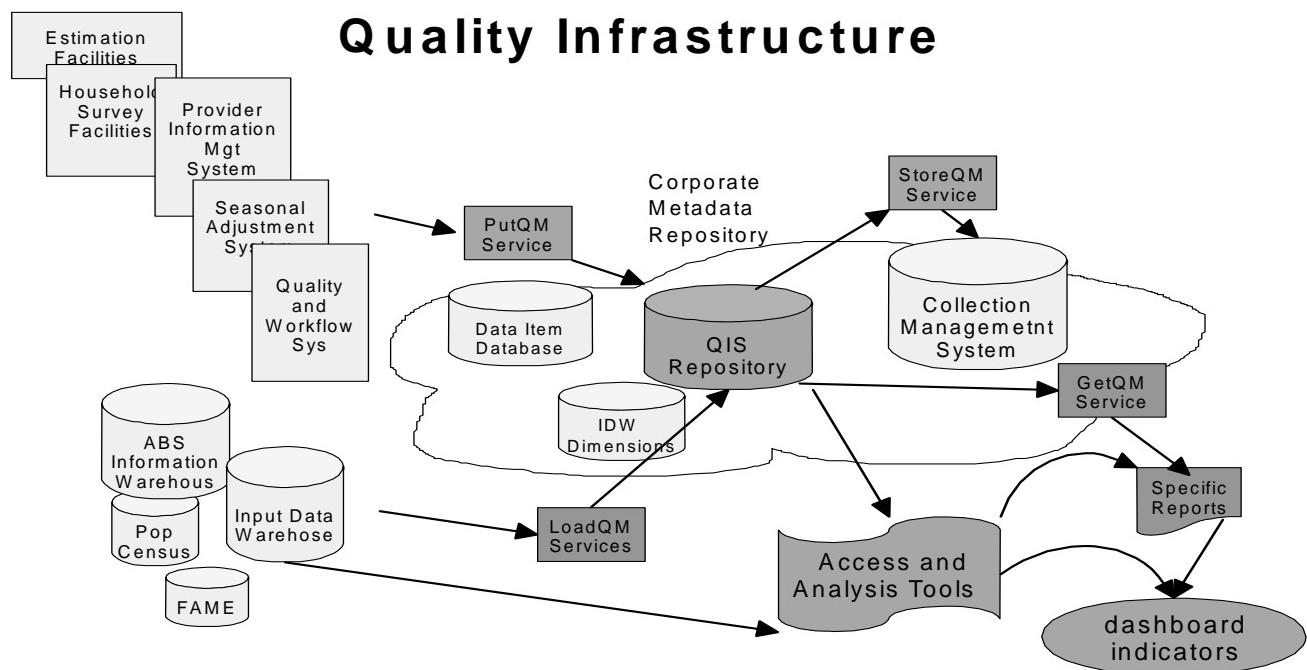
4. EXPANDING THE INFORMATION AVAILABLE EXTERNALLY

Over its history, the ABS, like other statistical offices, has developed an extensive range of ways to inform users about the quality of published data. See Lee and Allen (2001) for a summary description of the various approaches adopted by the ABS for making quality visible through hard copy publication. The quality assurance processes are well embedded in the way we conduct our business. However, as noted in section 1, more and more of users receive their data in electronic form only, posing new challenges for the ABS for developing quality assurance procedures for electronic outputs. Trewin (2002) describes the ABS response to these challenges, which includes, in particular, a rigorous clearance process for electronic outputs which are centrally stored in the data warehouse and disseminated centrally; and placing concepts, sources and methods publications on the website and providing electronic copies of an assortment of information and working papers to draw users' attention to particular issues that affect the interpretation or quality of the data.

The ABS recognises that further development of presentation standards and associated production tools is required for websites, datacubes, supertables, etc. Research was undertaken on how this may be done and a corporate wide approach is emerging. Standards and tools need to be established to enable author areas use the materials available from corporate metadata repository (such as the quality infrastructure system and collection management system) to include the information about non- sampling and sampling errors in the electronic output.

The ABS already has a repository of metadata about collection, including methodology and quality assessments. This repository is known as the Collection Management System. We investigated the possibility of producing a view for the Directory of Statistical Sources (an electronic reference which provides a comprehensive information about each statistical source) which assembles text material sourced from the Collection Management System. Our research has

produced a prototype which uses the six dimensions of the quality framework to present information in a "Quality Declaration". The view can be set up as hyperlinks within electronic publication.

Hyperlinks are also useful for linking components of an electronic products (such as table headings, variable descriptions, and annotations for sampling errors) to the relevant concepts, sources and methods materials, for example, relevant parts of the Directory of Statistical Sources. This approach, which the ABS has successfully used for a recently upgraded format for time series spreadsheets disseminated to users, illustrates how the explanatory notes in an electronic publication could be made more accessible at the point at which they are required.

Author areas at the moment must load the suitable material on the Collection Management System before these materials could be used for release. This is burdensome as loading and checking the information would demand resource at a critical point in the processing cycle. Development of corporate facilities for signing off electronic dissemination is an ABS priority.

5. CONCLUSION

In this paper I have described the plans and activities undertaken by the ABS to make the quality of statistical information more visible to users of statistics and ourselves. We have adopted an "inside out" approach for implementing changes. We started with influencing the conceptual understanding of ABS staff, and by stages seek to influence the language and culture of external users. We educate our staff on the role of information about quality in assessing fitness for purpose of particular datasets, and by stages extend the education and exhortation to external users and other government agencies involved in collecting and compiling statistics. We use information about quality to sharpen our focus on quality and improve our statistical processes, so that we could provide leadership for users to improve their understanding of quality. We build infrastructure to improve the useability of our toolsets for producing and disseminating quality information, in order to expand the quantum of quality information externally available.

We get the best results when changes are implemented in a wholistic manner. The MQV program addresses simultaneously issues relating to concepts and methodology, technology, business processes, people's skills and culture. Without this wholistic approach, it is difficult to ensure that we achieve the intended outcomes even when individual projects may have delivered the outputs expected.

Making quality visible and improving the quality of the statistical outputs itself, compete for the time and energies of our staff. Our approach is, as far as possible, to integrate the MQV activities with other corporate initiatives aimed at improving the effectiveness and efficiency of the statistical program. For example, the alignment of the work on quality standards with the extension of national statistical service, developing the electronic explanatory notes as the ABS expands its electronic dissemination capability; changing the way we produce, store, and disseminate quality information while the organisation is re-engineering its statistical processes.

Our resolve to improve the relevance, accessibility, timeliness, accuracy, coherence and interpretability of information about quality has been reinforced by the focus and attention that other statistical agencies and international organisations are giving to this issue.

As expectations from users increase, we have to accept that "the bar is rising" (Trewin, 2002). It is even more important for national statistical offices to manage the expectations of their users. Some further challenges for the ABS include:

- extending the national statistical service by working with other government agencies to make more statistical data available from administrative systems. Not only will we need to make quality visible for ABS data, the initiative has to extend to non-ABS data of importance to society;
- growing demand for electronically disseminated information, especially the expanded use of web service and other computer based analysis of data, will make it even harder for the ABS to make quality visible to the final user of the statistical analysis; and

- expanded range of statistical products and providers in the 'market of statistical information' will demand a greater capability for national statistical office to appropriately brand its products and make the value and cost benefits of quality more visible to their customers.

ACKNOWLEDGEMENT

The Making Quality Visible strategy described in this paper is based on an internal ABS paper prepared by Geoff Lee (2002). Many staff of the ABS, notably Bill Allen, Gemma Van Halderen, Paul Sutcliffe and Tenniel Guiver, have contributed to the Making Quality Visible project and their ideas are incorporated in this paper.

REFERENCES

Allen, B (2001), "Qualifying Quality - Issues of Presentation and Education", Proceedings of Statistics Canada Symposium 2001, Statistics Canada.

Australian Bureau of Statistics (2001), "National Statistical Service Best Practice Guidelines", unpublished report, Australian Bureau of Statistics.

Brackstone, G. (1999), "Managing Data Quality in a Statistical Agency", Survey Methodology, 25(2), pp. 139-149.

Carson, C. (2000), "Toward a Framework for Assessing Data Quality", The Proceedings of Statitical Quality Seminar, Jeju, Korea, Korea National Statistical Office and International Monetary Fund, pp. 33-106.

Fellegi, I. (1996), "Characteristics of an Effective Statistical System", International Statistical Review, 64, pp.165-197.

Eurostat (2001), Summary Report from the Leadership Group (LEG) on Quality, Luxembourg, 31 July 2001.

Lee, G. (2002), "Making Data Quality Visible - a focus for 2002/03", unpublished report, Australian Bureau of Statistics.

Lee, G. and Allen, B. (2001), "Educated Use of Information about Data Quality", Proceedings of the International Statisitcal Institute, Invited Papers, Seoul, Ausust 22-29, 2001,

Organisation for Economic Cooperation and Development (2003), Quality Framework and Guidelines for OECD Statistical Activities, Version 2003/1, STD/QFS(2003)1, OECD..

Trewin, D. (2002), "The Importance of a Quality Culture", Survey Methodology, 28(2), pp.125-133.

Zarb, J. (2001), "Quality Measures for Systems of Economic Accounts", unpublished paper presented to the Methodological Advisory Committee, 22 June 2001, Australian Bureau of Statistics.